

INTERNATIONAL  
STANDARD

ISO  
11940

First edition  
1998-06-01

---

---

**Information and documentation —  
Transliteration of Thai**

*Information et documentation — Translittération du thaï*

STANDARDSISO.COM : Click to view the full PDF of ISO 11940:1998



Reference number  
ISO 11940:1998(E)

## Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

International Standard ISO 11940 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Subcommittee SC 2, *Conversion of written languages*.

Annex A forms an integral part of this International Standard. Annexes B and C are for information only.

© ISO 1998

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

International Organization for Standardization  
Case postale 56 • CH-1211 Genève 20 • Switzerland  
Internet central@iso.ch  
X.400 c=ch; a=400net; p=iso; o=isocs; s=central

Printed in Switzerland

# Information and documentation — Transliteration of Thai

## 1 Scope

This International Standard establishes a system for the transliteration<sup>1)</sup> of Thai characters into Roman characters. One set of rules is provided for a completely reversible stringent conversion where principles of transliteration are applied without exception.

This Romanized transliteration system is a means of converting the Thai writing system into a Roman alphabet writing system. Since there are fewer Roman characters than Thai characters, diacritical marks, punctuation marks, and a combination of two Roman characters, or a digraph, are needed to represent one Thai character. The aim of this system is to provide a means of international communication of written messages in a form which permits the automatic transmission or reconstitution of these by men or machines. This system of conversion is intended to provide complete and unambiguous reversibility. It is possible that the results obtained from this system will not provide for correct pronunciation of the original Thai text. However, they will serve as a means of finding the Thai graphisms automatically and thus to allow anyone with a knowledge of Thai to pronounce the Thai text correctly. An attempt to transpose a preposed vowel after an initial consonant may comfort those who are used to the Roman writing system where a vowel always follows a consonant to form a syllable, unless it is a syllable without an initial. It is best to leave the preposed vowel in the Thai transliteration system where it is, because a transliteration system is not a transcription system.

Capital letters are reserved for writing the initial in proper nouns. It is unwise to make use of capital letters in the transliteration system for certain characters, otherwise common nouns and proper nouns cannot be distinguished from one another.

This system is based on phonetic logic in the selection of transliterated representations, as much as possible, in order to lessen the burden on those who wish to learn the system.

## 2 Normative reference

The following standard contains provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the edition indicated was valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent edition of the standard indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 10646-1:1993, *Information Technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*.

## 3 Definitions

For the purposes of this International Standard, the following definitions apply.

**3.1 characters:** Alphabetic letters, digits, special markers, and other markers.

**3.2 Thai characters:** Thai alphabetic letters, Thai digits, and special markers which are divided into the following:

1) Definitions, explanations for terms used (transliteration, transcription, etc.) appear in annex A.

3.2.1 Thai alphabetic letters:

3.2.1.1 Consonants:

ก ข ฃ ค ฅ ฉ ง จ ฉ ช ซ ฌ ญ ฎ ฏ ฐ ฑ ฒ ณ  
 ด ต ถ ท ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ฤ ล ฦ  
 ว ศ ษ ส ห ฬ อ ฮ

3.2.1.2 Vowels

3.2.1.2.1 Vowels written on the same level as consonants:

ะ ำ ำ ะ ั ็ ็ ็ ็

3.2.1.2.2 Vowels written above the line (upper level characters):

ๅ ๆ ็ ๘ ๙ ๐

3.2.1.2.3 Vowels written below the line (lower level characters):

๑ ๒ ๓

3.2.1.3 Tone Marks and Cancellation Mark written in the uppermost level<sup>2)</sup> (uppermost level characters):

๔ ๕ ๖ ๗ ๘

3.2.2 Thai Digits:

๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙

3.2.3 Special Markers:

๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙

**3.3 special symbols:** Refers to the four symbols used to differentiate one Roman character that represents the same sound converted from different Thai characters:

- 3.3.1 macron —
- 3.3.2 macron below \_
- 3.3.3 dot below .
- 3.3.4 horn ’

2) If a tone mark or a cancellation mark is directly over a consonant without an upper level character, it is adjusted to an upper level in typesetting and displaying on the computer screen for neatness of appearance.

## 4 Arrangement of character sequence

4.1 The conversion is made character by character from left to right.

4.2 A consonant can have either an upper level or a lower level character and/or an uppermost level character; the sequence of the characters must be as follows:

4.2.1 A consonant with either an upper level character or an uppermost level character, e.g.

$\overset{\text{a}}{\text{b}} = \text{b}^{\text{a}}$ , or  $\overset{\text{v}}{\text{b}} = \text{b}^{\text{v}}$ :

- Type the consonant first; then
- Type the upper level character or the uppermost level character.

4.2.2 A consonant with both an upper level character and an uppermost level character, e.g.

$\overset{\text{a}}{\text{b}}^{\text{v}} = \text{b}^{\text{v}\text{a}}$ :

- Type the consonant first; then
- Type the uppermost level character, and
- Type the upper level character.

4.2.3 A consonant with a lower level character, e.g.

$\underset{\text{v}}{\text{b}} = \text{b}_v$ :

- Type the consonant first; then
- Type the lower level character.

4.2.4 A consonant with an uppermost level character and a lower level character, e.g.

$\overset{\text{v}}{\underset{\text{a}}{\text{b}}} = \text{b}^{\text{v}}_a$ :

- Type the consonant first; then
- Type the uppermost level character, and
- Type the lower level character.

It must be noted that both an upper level character and a lower level character cannot concurrently appear with the same consonant.

## 5 Typing sequence

5.1 If a Roman character contains one special symbol, type the character before the symbol.

5.2 If a Roman character contains two special symbols, type the character first, then the macron, and the dot below or the horn respectively.

## 6 Romanization table

No.	Thai character	Code position of Thai character in ISO/IEC 10646-1	Romanized character	Code position(s) of romanized characters in ISO/IEC 10646-1	Comments
1	ก	0E01	k	006B	Ko Kai
2	ข	0E02	kh	006B+0304+0068	Kho Khai
3	ฃ	0E03	kh̄	006B+0304+0323+0068	Kho Khuat
4	ค	0E04	kh	006B+0068	Kho Khwai
5	ค	0E05	k'h	006B+031B+0068	Kho Khon
6	ฆ	0E06	kh̄	006B+0304+0068	Kho Rakhang
7	ง	0E07	ng	006E+0067	Ngo Ngu
8	จ	0E08	c	0063	Cho Chan
9	ฉ	0E09	ch̄	0063+0304+0068	Cho Ching
10	ช	0E0A	ch	0063+0068	Cho Chang
11	ซ	0E0B	s	0074	So So
12	ฌ	0E0C	ch̄	0063+0323+0068	Cho Choe
13	ญ	0E0D	y	0079+0323	Yo Ying
14	ฎ	0E0E	ḍ	0064+0323	Do Chada
15	ฏ	0E0F	ṭ	0074+0323	To Patak
16	ฐ	0E10	th̄	0074+0304+0323+0068	Tho Than
17	ฑ	0E11	th	0074+0331+0068	Tho Nangmontho
18	ฒ	0E12	t'h	0074+031B+0068	Tho Phuthao
19	ณ	0E13	ṇ	006E+0323	No Nen
20	ด	0E14	d	0064	Do Dek

No.	Thai character	Code position of Thai character in ISO/IEC 10646-1	Romanized character	Code position(s) of romanized characters in ISO/IEC 10646-1	Comments
21	ต	0E15	t	0074	To Tao
22	ถ	0E16	th	0074+0304+0068	Tho Thung
23	ท	0E17	th	0074+0068	Tho Thahan
24	ธ	0E18	th	0074+0323+0068	Tho Thong
25	น	0E19	n	006E	No Nu
26	บ	0E1A	b	0062	Bo Baimai
27	ป	0E1B	p	0070	Po Pla
28	ผ	0E1C	p̄h	0070+0304+0068	Pho Phung
29	ฝ	0E1D	f̄	0066+0304	Fo Fa
30	พ	0E1E	ph	0070+0068	Pho Phan
31	ฟ	0E1F	f	0066	Fo Fan
32	ภ	0E20	ph	0070+0323+0068	Pho Samphao
33	ม	0E21	m	006D	Mo Ma
34	ย	0E22	y	0079	Yo Yak
35	ร	0E23	r	0072	Ro Rua
36	ฤ	0E24	v	0076	Ru
37	ล	0E25	l	006C	Lo Ling
38	ฬ	0E26	ł	0142	Lu
39	ว	0E27	w	0077	Wo Waen
40	ศ	0E28	s̄	0073+0304+0323	So Sala
41	ษ	0E29	s̄	0073+0304+031B	So Rusi

No.	Thai character	Code position of Thai character in ISO/IEC 10646-1	Romanized character	Code position(s) of romanized characters in ISO/IEC 10646-1	Comments
42	ส	0E2A	s̄	0073+0304	So Sua
43	ห	0E2B	h̄	0068+0304	Ho Hip
44	ฬ	0E2C	l̄	006C+0323	Lo Chula
45	อ	0E2D	x	0078	O Ang
46	ฮ	0E2E	h̄	0068+0323	Ho Nokhuk
47	ะ	0E30	a	0061	Sara A
48	ั	0E31	ā	0061+0323	Mai Han-Akat
49	า	0E32	ā	0061+0304	Sara Aa
50	ำ	0E33	ā̄	00E5	Sara Am
51	ิ	0E34	i	0069	Sara I
52	ี	0E35	ī	012B	Sara Ii
53	ุ	0E36	u	0075+0323	Sara Ue
54	ู	0E37	ū	0075+0304+0323	Sara Uee
55	ุ	0E38	u	0075	Sara U
56	ู	0E39	ū	0075+0304	Sara Uu
57	เ	0E40	e	0065	Sara E
58	แ	0E41	æ	00E6	Sara Ae
59	โ	0E42	o	006F	Sara O
60	ไ	0E43	i	0131	Sara Ai Maimuan
61	ใ	0E44	i	0069+0323	Sara Ai Maimalai
62	ย	0E45	i	0268	Lakkhangyao

No.	Thai character	Code position of Thai character in ISO/IEC 10646-1	Romanized character	Code position(s) of romanized characters in ISO/IEC 10646-1	Comments
63	ิ	0E48	˘	0300	Mai Ek
64	ุ	0E49	ˆ	0302	Mai Tho
65	ึ	0E4A	ˇ	0301	Mai Tri
66	๓	0E4B	˘˘	030C	Mai Chattawa
67	๓	0E47	˘	0306	“Maitaikhu” shortens a long vowel syllable, when a short vowel symbol ๓ does not exist.
68	๓	0E4C	˘	0312	“Thanthakhat” silences a character or a group of characters when it appears above that character, which may be the last character of the group, such as สังข์ พักตร์ ชอล์ก ลักขมณ.
69	๓	0E46	«	00AB	“Maiyamok” indicates a repetition of a word, a group of words or a phrase, or sometimes even a sentence.
70	๓	0E2F	≠	01C2	“Paiyanoi” indicates that the word has been shortened as in กรู่งเทพ๓. It can combine with ๓ as ๓๓๓ which is equivalent to et cetera.
71	๓	0E4E	˘	007E	“Yamakkan” is written over the first member of a consonant cluster in old texts to indicate that the two consonants must be pronounced as a consonant cluster.

No.	Thai character	Code position of Thai character in ISO/IEC 10646-1	Romanized character	Code position(s) of romanized characters in ISO/IEC 10646-1	Comments
72	๐	0E4F	§	00A7	“Takai” or “Fongman” is used at the beginning of a stanza in poetry and also the beginning of a paragraph in old texts. ๐ “Fongman Fannu” occurs in older texts and can be considered a variant of ๐.
73	๑	-		01C0	“Angkhandiao” is used at the end of a sentence, and at the end of a stanza in old texts. It is also used as a middle symbol in writing dates in the lunar calendar, for example ๑๑๑ meaning Wednesday, the 15th day (full moon) of the sixth lunar month. (Note: it looks very similar to ๑, “Paiannoï” in # 70.)
74	๒	0E5A		01C1	“Angkhankhu” is used to terminate a paragraph in old texts. It can also combine with ๒, “Khomut” in # 75 as ๒๒ to terminate a story.
75	๒	0E5B	»	00BB	“Khomut” is written at the end of a story to indicate termination in old texts.

No.	Thai character	Code position of Thai character in ISO/IEC 10646-1	Romanized character	Code position(s) of romanized characters in ISO/IEC 10646-1	Comments
76	◌̇	0E4D	◌̇	030A	“Nikkhahit” is written above the rightmost shoulder of a character to represent an ‘ang’ or ‘an’ sound, such as  pronounced [sǎŋkhǎŋ], and above the rightmost shoulder of a character and a vowel to represent an ‘m’ sound, such as  pronounced [chumnum].
77	◌̣	0E3A	◌̣	0325	“Phinthu” is written under a single consonant to signify a final consonant or under the first member of double consonants to signify a consonant cluster in Pali texts.
78	๐	0E50	0	0030	Zero
79	๑	0E51	1	0031	One
80	๒	0E52	2	0032	Two
81	๓	0E53	3	0033	Three
82	๔	0E54	4	0034	Four
83	๕	0E55	5	0035	Five
84	๖	0E56	6	0036	Six
85	๗	0E57	7	0037	Seven
86	๘	0E58	8	0038	Eight
87	๙	0E59	9	0039	Nine

## Annex A (normative)

### Principles of conversion

#### A.1 Standards on conversion of systems of writing

This International Standard is one of a series of International Standards, dealing with the conversion of systems of writing. The aim of this International Standard and others in the series is to provide a means for international communication of written messages in a form which permits the automatic transmission and reconstitution of these by men or machines. The system of conversion, in this case, must be univocal and entirely reversible.

This means that no consideration should be given to phonetic and aesthetic matters nor to certain national customs: all these considerations are, indeed, ignored by the machine performing the function.

The adoption of this International Standard for international communication leaves every country free to adopt for its own use a national standard which may be different, on condition that it be compatible with the International Standard. The system proposed herein should make this possible and be acceptable to international use if the graphisms it creates are such that they may be converted automatically into the graphisms used in any strict national system.

This International Standard may be used by anyone who has a clear understanding of the system and is certain that it can be applied without ambiguity. The result obtained will not give a correct pronunciation of the original text in a person's own language but it will serve as a means of finding automatically the original graphism and thus allow anyone who has a knowledge of the original language to pronounce it correctly. Similarly, one can only pronounce correctly a text written in, for example, English or Polish, if one has a knowledge of English or Polish.

The adoption of national standards compatible with this International Standard will permit the representation, in an international publication, of the morphemes of each language according to the customs of the country where it is spoken. It will be possible to simplify this representation in order to take into account the number of the character sets available on different kinds of machines.

#### A.2 General principles of conversion of writing systems

##### A.2.1 Definition and methods

**A.2.1.1** The words in a language, which are written according to a given script (the converted system), sometimes have to be rendered according to a different system (the conversion system) normally used for a different language. The procedure is often used for historical or geographical texts, cartographical documents and in particular bibliographical work where characters must be converted from different writing systems into a single alphabet to allow for alphabetical intercalation in bibliographies, catalogues, indexes, toponymic lists, etc. It is indispensable in that it permits the univocal transmission of a written message between two countries using different writing systems or exchanging a message the writing of which is different from their own.

It thereby permits transmission by manual, mechanical as well as electronic means.

The two basic methods of conversion of a system of writing are transliteration and transcription.

**A.2.1.2 Transliteration** is the process which consists of representing the characters<sup>3)</sup> of an alphabetical or syllabic system of writing by the characters of a conversion alphabet, this being the easiest way to ensure the complete and unambiguous reversibility of the conversion alphabet in the converted system.

---

3) A character is an element of an alphabetical or other type of writing system that graphically represents a phoneme, a syllable, a word or even a prosodical characteristic of a given language. It is used either alone (e.g. a letter, a syllabic sign, an ideographical character, a digit, a punctuation mark) or in combination (e.g. an accent, a diacritical mark). A letter having an accent or a diacritical mark, for example â, è, ö, is therefore a character in the same way as a basic letter.

In exceptional cases, e.g. when the number of characters used in the conversion system is smaller than the number of characters of the converted system, it is necessary to use digraphs or diacritical marks. In this case, one must avoid as far as possible arbitrary choice and the use of purely conventional marks, and try to maintain a certain phonetic logic in order to give the system a wide acceptance.

However, it must be accepted that the graphism obtained may not always be correctly pronounced according to the phonetic habits of the language(or of all the languages) which usually use(s) the conversion alphabet. On the other hand this graphism must be such that the reader who has a knowledge of the converted language may mentally restore unequivocally the original graphism and thus pronounce it.

**A.2.1.3 Retransliteration** is the process whereby the characters of a conversion alphabet are transformed back into those of the converted writing system. It is the exact opposite of the transliteration process in that the rules of a transliteration system are applied in reverse in order to reconvert the transliterated word to its original form.

**A.2.1.4 Transcription** is the process whereby the sounds of a given language are noted by the system of signs of a conversion language.

A transcription system is of necessity based on the orthographical conventions of the conversion language. Transcription is not strictly reversible.

Transcription may be used for the conversion of all writing systems. It is the only method that can be used for systems that are not entirely alphabetical or syllabic and for all ideophonographical systems of writing like Chinese.

**A.2.1.5** To carry out **romanization**, the conversion of non-Latin writing systems to the Latin alphabet, either transliteration or transcription or a combination of the two may be used depending on the nature of the converted system.

**A.2.2** A conversion system proposed for international use may call for compromise and the sacrifice of certain national customs. It is therefore necessary for each community of users to accept concessions, fully abstaining in every case from imposing as a matter of course solutions that are actually justified only by national practice (for example as regards pronunciation, orthography, etc.).

When a country uses two systems univocally convertible one into the other to write its own language, the system of transliteration thus implemented must be taken a priori as a basis for the international standardized system, as far as it is compatible with the other principles exposed hereafter.

**A.2.3** When necessary, the conversion systems should specify an equivalent for each character, not only the letters but also the punctuation marks, numbers, etc. They should similarly take into account the arrangement of the sequence of characters that make up the text, for example the direction of the script, and specify the way of distinguishing words and of using separation signs, following as closely as possible the customs of the language(s) which use the converted writing system.

**A.2.4** When romanizing a script which does not have upper-case characters, it is usual to capitalize some words, following national usage.

### **A.3 Principles of conversion for alphabetical writing systems**

**A.3.1** The conversion may be made at various levels

The first level is that of completely reversible **stringent transliteration** which is necessary to attain, in full, the aim given in clause A.2.1.2. This conversion applies all principles of transliteration without exception. However, whenever it is useful to distinguish the end or beginning of a syllable (a morpheme or a word) variants may be used. The conventional systems of stringent transliteration shall be applied as such without any change to meet national or regional customs as regards pronunciation or orthography. They permit the univocal international transmission of messages by mechanical or electronic means.

To permit an international unequivocal communication, International Standards on transliteration shall apply by priority the principle of stringent conversion. They can then be used as a basis for the establishment of rules for simplified conversion and for preparation of national standards.

The second level is that of **simplified conversion**. The simplification may be made necessary, for example, by the use of machines that do not accept all the alphabet characters required for stringent conversion. The method of conversion may allow national or regional variants, which may not permit complete reversibility. The simplified conversion may be the subject of International Standards or agreements.

The third level is that of **popular conversion** which, for example, should enable the same foreign names to be written in a uniform manner in the newspapers of a given country. It is obliged to take into account phonetic or graphic practices and therefore can only be national.

**A.3.2** In cases where the same characters appear in one alphabet used with some differences by different languages, these characters would be transliterated in the same way, irrespective of the language they belong to.

**A.3.3** If the converted alphabet gives a different form to the same character according to its place in the word (as is the case for example in the Arabic, Hebrew and Greek alphabets), the conversion alphabet will use only one character of constant form.

STANDARDSISO.COM : Click to view the full PDF of ISO 11940:1998